

530,953

APR 2005

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
22 April 2004 (22.04.2004)

PCT

(10) International Publication Number  
**WO 2004/034276 A1**

- (51) International Patent Classification<sup>7</sup>: **G06F 17/00**
- (21) International Application Number:  
PCT/SG2003/000233
- (22) International Filing Date:  
25 September 2003 (25.09.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
200206227-1 11 October 2002 (11.10.2002) SG
- (71) Applicant (*for all designated States except US*): MAT-SUSHITA ELECTRIC INDUSTRIAL CO. LTD. [JP/JP]; 1006, Oaza Kadoma, Kadoma-shi, Osaka 571-8501 (JP).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): TAN, Jek-Thoon, [SG/SG]; Blk 56, #02-30, Choa Chu Kang North 6, Yew Mei Green, 689577 (SG). SHEN, Sheng Mei, [SG/SG]; No. 20, #03-02, Choa Chu Kang Street 64, Windermere, 689093 (SG).
- (74) Agent: ELLA CHEONG MIRANDAH & SPRUSONS PTE. LTD., P.O. Box 1531, Robinson Road Post Office, 903031 (SG).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Declaration under Rule 4.17:**  
— *of inventorship (Rule 4.17(iv)) for US only*
- Published:**  
— *with international search report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: A METHOD AND APPARATUS FOR DELIVERING PROGRAMME-ASSOCIATED DATA TO GENERATE RELEVANT VISUAL DISPLAYS FOR AUDIO CONTENTS

(57) Abstract: An MPEG audio stream is transmitted together with an MPEG video stream. The audio stream contains an audio signal together with associated audio description data as ancillary data. The video stream contains a video signal together with video description data (e.g. video clips, stills, graphics, text etc) as private data, the video description data not necessarily having anything to do with the video data with which it is transmitted. At reception, the audio and video streams are decoded. The video description data is stored in a memory. The audio signal is played. The audio description data is used to select appropriate video description data for the particular audio signal from the memory or other storage, or from the current incoming video description data. This is then displayed as the audio signal is played.

WO 2004/034276 A1

**A METHOD AND APPARATUS FOR DELIVERING PROGRAMME-ASSOCIATED  
DATA TO GENERATE RELEVANT VISUAL DISPLAYS FOR AUDIO CONTENTS**

**TECHNICAL FIELD**

5

The present invention relates to the provision of an audio signal with an associated video signal. In particular, it relates to the use of audio description data, transmitted with an audio signal as part of an audio stream, to select an appropriate video signal to accompany the audio signal during playback.

10

**BACKGROUND TO THE INVENTION**

In digital music media and broadcast applications such as MP3 players and digital audio broadcast, the experience is usually solely audio. When listening to music, people usually tend only to listen, without watching anything. The audio programme is usually played without giving the listener any interesting visual display.

In some standards, ancillary data may be carried within an audio elementary stream for broadcast or storage in audio media. The most common use of ancillary data is programme-associated data, which is data intimately related to the audio signal. Examples of programme-associated data are programme related text, indication of speech or music, special commands to a receiver for synchronisation to the audio programme, and dynamic range control information. The programme-associated data may contain general information such as song title, singer and music company names. It gives relevant facts but is not useful beyond that.

In current digital TV developments, programme-associated data carrying textual and interactive services can be developed for the TV programmes. These solutions cover implementation details including protocols, common API languages, interfaces and recommendations. The programme-associated data are transmitted together with the video and audio content multiplexed within the digital programme or transport stream. In such implementations, relevant programme-associated data must be developed for each TV programme, and there must also be constant monitoring of the multiplexing process. Besides, this approach occupies transmission bandwidth.

35

Developing content for programme-associated data requires significant manpower resources. As a result, the cost of delivering such applications is high, especially when different contents have to be developed for different TV programmes. It would also be desired that such programme-associated data contents could be reused for different video, audio and TV programmes.

Other attempts have been made which involve displaying something sometimes during audio playback, in particular for karaoke.

Japanese patent publication No. JP10-124071 describes a hard disk drive provided with a music data storage part which stores music data on pieces of karaoke music and a music information database which stores information regarding albums containing these pieces of music. In the music data, a flag is provided showing whether or not the music is one contained in an album. A controller determines if a song is one for which the album information is available. During an interval for a song where the information is available, data on the album name and music are displayed as a still picture.

Japanese patent publication No. JP10-268880 describes a system to reduce the memory capacity needed to store respective image data, by displaying still picture data and moving picture data together according to specific reference data. Genre data in the header part of Karaoke music performance data is used to refer to a still image data table to select pieces of still image data to be displayed during the introduction, interlude and postlude of the song. The genre data is also used to refer to a moving image data table to select and display moving image data at times corresponding to text data.

According to patent publication JP2001-350482A Karaoke data can include time interval information indicating time bands of non-singing intervals. For a performance, this information is compared with presentation time information relating to a spot programme. The spot programme whose presentation time is closest to the non-singing interval time is displayed during that non-singing interval.

Japanese patent publication No. JP7-271,387 describes a recording medium which records audio and video information together so as to avoid a situation in which a

singer merely listens to the music and waits for the next step while a prelude and an interlude are being played by Karaoke singing equipment. A recording medium includes audio information for accompaniment music of a song and picture information for a picture displaying the text of the song. It also includes text picture information for a text picture other than the song text.

According to Japanese patent publication No. JP2001-350,482 Karaoke data can include time interval information indicating time bands of non-singing intervals. During playback, this information is compared with presentation time information relating to a spot programme. The spot programme whose presentation time is closest to the non-singing interval time is displayed during that non-singing interval.

#### SUMMARY OF THE INVENTION

The present invention aims to provide the possibility of generating exciting and interesting visual displays. It may be desired to generate changing visual content relevant to the audio programme, for example beautiful scenery for music and relevant visual objects for various theme music, songs or lyrics.

According to one aspect of the present invention, there is provided a method of providing an audio signal with an associated video signal, comprising the steps of:

decoding an encoded audio stream to provide an audio signal and audio description data; and

providing an associated first video signal at least part of whose content is selected according to said audio description data.

Preferably said providing step comprises:

using said audio description data to select visual description data appropriate to the content of said audio signal; and

constructing video content from said selected visual description data; and providing said first video signal including the constructed video content.

The method may further comprise the step of extracting said visual description data from a transport stream, for instance an MPEG stream containing audio, video and the visual description data.

According to a second aspect of the present invention, there is provided a method of delivering programme-associated data to generate relevant visual display for audio contents, said method comprising the steps of:

- 5        encoding an audio signal and audio description data associated therewith into an encoded audio stream;  
         encoding visual description data; and  
         combining said encoded audio stream and said visual description data.

The first and second aspects may be combined.

10

According to a third aspect of the present invention, there is provided apparatus for providing an audio signal with an associated video signal, comprising:

- audio decoding means for decoding an encoded audio stream to provide an audio signal and audio description data; and  
15        first video signal means for providing an associated first video signal at least part of whose content is selected according to said audio description data.

According to a fourth aspect of the present invention, there is provided a system for providing an audio signal with an associated video signal, comprising:

- 20        audio encoding means for encoding an audio signal and audio description data into an encoded audio stream  
         description data encoding means for encoding visual description data; and  
         combining means for combining said encoded audio stream and said visual description data.

25

The third and fourth aspects may be combined.

According to a fifth aspect of the present invention, there is provided a system for delivering programme-associated data to generate relevant visual display for audio contents, said system comprising:

- 30        audio encoding means for encoding an audio signal and audio description data associated therewith into an encoded audio stream;  
         video encoding means for encoding visual description data into an encoded video stream; and  
35        combining means for combining said encoded audio and video streams.

In any of the above aspects, said visual description data is capable of comprising one or more of the group comprising: video clips, still images, graphics and textual descriptions. Alternatively or additionally, said visual description data may be  
5 classified for use with at least one of: at least one style of audio content, at least one theme of audio content and at least one type of event for which it might be suitable.

Said audio description data may comprise data relating to at least one of the group comprising: singer identification, group identification, music company identification,  
10 service provider identification and karaoke text. Alternatively or additionally, said audio description data may comprise data relating to the style of said audio signal. Alternatively or additionally again, said audio description data may comprise data relating to the theme of audio signal. As another possibility, said audio description data may comprise data relating to the type of event for which said audio signal might  
15 be suitable.

The audio description data may be within frames of said encoded audio stream, which frames also containing said audio signal. The encoded audio stream may be an MPEG audio stream. Where both occur, then said audio description data may be  
20 ancillary data within said MPEG audio stream.

In another aspect of the invention, any of the above apparatus or systems is operable according to any of the above methods.

25 Thus the invention provides an audio signal with an associated video signal. In particular, it provides an audio description data, transmitted with an audio signal as part of an audio stream, to select an appropriate video signal to accompany the audio signal.

30 This invention provides an effective means of adding further information relevant to the audio programme. It creates an option for the content provider to insert or modify relevant information describing the audio content for generating relevant visual content prior distributing or broadcasting. The programme-associated data, which may be carried in the ancillary data section of the audio elementary stream, provides a general

description of the preferred classification or categories for use by the decoder to generate relevant visual display and interactive applications.

It may be desirable to insert programme-associated data to generate relevant, exciting and interesting visual displays for a listener, for example sports scenes or still pictures for sports related songs or lyrics. To generate such visual displays, a method of encoding and inserting the programme-associated data in the audio elementary streams, as well as a technique of decoding, interpreting and generating the visual display is provided. This invention provides an effective means of adding further information relevant to the audio programme. The programme-associated data carried in the ancillary data section of the audio elementary stream shall provide general description of the preferred classification or categories for use by the decoder to generate relevant visual display and interactive applications.

In one aspect, an MPEG audio stream is transmitted together with an MPEG video stream. The audio stream contains an audio signal together with associated audio description data as ancillary data. The video stream contains a video signal together with video description data (e.g. video clips, stills, graphics, text etc) as private data, the video description data not necessarily having anything to do with the video data with which it is transmitted. At reception, the audio and video streams are decoded. The video description data is stored in a memory. The audio signal is played. The audio description data is used to select appropriate video description data for the particular audio signal from the memory or other storage, or from the current incoming video description data. This is then displayed as the audio signal is played.

#### INTRODUCTION TO THE DRAWINGS

The present invention will now be further described by way of non-limitative example with reference to the accompanying drawings, in which:-

Figure 1 is a block diagram of encoding audio and video description data;

Figure 2 is a block diagram of a receiver of one embodiment of the invention; and

Figure 3 is a schematic view of what happens at a receiver embodying the present invention;

### DETAILED DESCRIPTION

5 In this invention, programme-associated data describing an audio content is used as a basis to generate a visual display for a listener, for example: short video clips, scenes, images, advertisements, graphics, textual and interactive contents on festive events for songs or lyrics related to special occasions, where the visual display is relevant to the audio content. Methods of encoding and inserting the programme-associated data in audio elementary streams are used to generate such visual displays.

10

The programme-associated data is used to generate visual display relevant to the audio content. It can be distinctly categorised into two types of data: (i) audio description data for describing the audio content and (ii) visual description data for generating the visual display. The visual description data need not be developed for  
15 specific audio programme or audio description data.

#### (i) audio description data

20 Audio description data gives general descriptions of the audio content such as the music theme, the relevant keyword for the song lyrics, titles, singer or company names, as well as the style of the music. The audio description data can be inserted in each audio frame or at various audio frames throughout the music or song duration, thus enabling different descriptions to be inserted at different sections of the audio programme.

25

#### (ii) visual description data

The visual description data may contain short video clips, still images, graphics and textual descriptions, as well as data enabling interactive applications. The visual  
30 description data can be encoded separately from the audio description data and is delivered to the receiver as private data, residing in private tables of the transport or programme streams. The visual description data need not be developed for specific audio programme or audio description data. It can be developed for specific audio "style", "theme", "events", and can also contain relevant advertising and interactive  
35 information.



Figure 1 is a block diagram of an encoding process for audio and visual description data according to an embodiment of the present invention.

- 5 An audio source 12 provides an audio signal 14 to an audio encoder 16, which encodes it into suitable audio elementary streams 18 for storing in a storage media 20, such as a set of hard discs.

10 An audio description data encoder 22 is a content creation tool for developing audio description data, such as general descriptions of the audio content. It is user operable or can work automatically, for example by analysing the musical and/or text content of the audio elementary streams (the tempo of music can for example be analysed to provide relevant information). The audio description data encoder 22 retrieves audio elementary streams from the storage media 20 and inserts the audio description data  
15 it creates into the ancillary data section within each frame of the audio elementary streams. After editing or inserting, the audio elementary stream containing the audio description data 24 is stored back in the storage media 20 for distribution or broadcast. The audio description data encoder 22 also produces identification and clock reference data 26 associated with the audio elementary stream containing the  
20 audio description data 24, and also stores these in the audio elementary stream.

A video/image source 28 provides a video/image signal 30 to a video/image encoder 32, which encodes it into a suitable data format 34 for storing in a storage media 36. Other data media 38 may also contribute suitable visual data 40 such as textual and  
25 graphics data. Archives of video clips, images, graphics and textual data 42 from the storage media 36 are supplied to and used by a visual description data encoder 44 for developing the visual content. The way this is done is platform dependent. For video clips they could be stored as MPEG-1/MPEG-2 or any one of a number of video formats that are supported. For graphics, they could be provided and stored as  
30 MPEG-4 or MPEG-7 description language or Java or such like. For text it could be provided and stored in unicode. For any of these, the definitions could even be proprietary.

The visual description data encoder 44 is a content creation tool for developing visual  
35 description data 46. The visual description data 46 is stored in a storage media 48 for

distribution or broadcast. The visual description data 46 may be developed independently from the audio content. However, for applications where the visual description data 46 is intended to be executed together with associated audio description data, the identification code and clock reference 26 from audio description data encoder 22 are used to synchronise the decoding of the visual description data. For this, they are included in private defined descriptors which are embedded in the private sections carrying the visual description data.

During broadcast, whether by cable, optical or wireless transmission and whether as television or internet, audio elementary streams (including the audio description data) from audio storage media 20 are multiplexed with the visual description data as private data from video storage media 36 and video elementary streams (for instance containing a video) to form a transport stream. This is then channel coded and modulated to transmission.

Figure 2 is a block diagram of a receiver constructed in accordance with another embodiment of the invention for digital TV reception. An RF input signal 50 is received and passed on to a front-end 52 controlled to tune in the correct TV channel. The front-end 52 demodulates and channel decodes the RF input signal 50 to produce a transport stream 54.

A transport decoder 56 extracts a private section table from the transport stream 54 by identifying a unique 13-bit PID that contains the visual description data. The visual description data is channelled through the decoder's data bus 58 to be stored in a cyclic buffer 60. At the same time the transport decoder 56 also filters the audio elementary stream 62 and video elementary streams 64 to an MPEG audio decoder 66 and MPEG video decoder 68 respectively, from the transport stream 54.

The PID (Program Identification) is unique for each stream and is used to extract the audio stream, the video stream and the private section data containing the visual description data.

The MPEG audio decoder 64 decodes the audio elementary stream 62 to produce the decoded digital audio signal 70. The decoded digital audio signal 70 is sent to an audio encoder 72 to produce an analogue audio output signal 74. The ancillary data

containing the audio description data in the audio elementary stream is filtered and stored in a cyclic buffer 76 via the audio decoder's data bus 78.

5 The MPEG video decoder 68 decodes the video elementary stream 64 to produce the decoded digital video signal 80. The decoded digital video signal 80 is sent to a graphics processor and video encoder 82 to produce the video output signal 84.

10 The receiver host microprocessor 86 controls the front-end 52 to tune in the correct TV channel via an I<sup>2</sup>C bus 88. It also retrieves the visual description data from the cyclic buffer 60 through the transport decoder's data buses 58, 90. The visual description data is stored in a memory system 92 via the host data bus 94. The visual description data may also be downloaded from external devices such as PCs or other storage media via an external data bus 96 and interface 98.

15 The microprocessor 86 also reads the filtered audio description data from the cyclic buffer 76 via the audio decoder's data buses 78, 100. From the audio description data, it uses cognitive and search engines to select the best-fit visual description data from the system memory 92. The general steps used in selecting the best-fit may be as follows:

- 20 i. retrieve audio description data from the audio elementary stream. This is identified by the "audio\_description\_identification" value (described later);
- ii. retrieve the "description\_data\_type" value (described later) to determine the type of data that follows;
- 25 iii. if the value of "description\_data\_type" is between 1 and 15, retrieve the "user\_data\_code" (Unicoded text) (described later) that describes the respective type of information. This information is used as the search criteria;
- iv. if the value of "description\_data\_type" is any of 16, 17 and 18, retrieve the "description\_data\_code" (described later) to determine the search criteria. The "description\_data\_code" follows the definitions described in Tables 5, 6 and 7  
30 (appearing later) for "description\_data\_type" values of 16, 17 and 18, respectively;
- v. search the visual description database of memory 92 for best matches based on the search criteria. The database contains the visual description data files, stored in directories with filenames organised to allow the use of an effective  
35 search algorithm.

The operation of the MPEG video decoder 68 is also controlled by the microprocessor 86, via the decoder's data bus 102.

5 The graphics processor and video encoder module 82 has a graphics generation engine for overlaying textual and graphics, as well as performing mixing and alpha scaling on the decoded video. The operation of the graphics processor is controlled by the microprocessor 86 via the processor's data bus 104. Selected best-fit visual description data from the system memory 92 is processed under the control of the  
10 microprocessor 86 to generate the visual display using the features and capabilities of the graphics processor. It is then output as the sole video output signal or superimposed on the video signal resulting from the video elementary stream.

Thus, in use, the receiver extracts the private data containing the visual description  
15 data and stores in its memory system. When an audio programme is played (even at a later time), the receiver extracts the audio description data and uses that to search its memory system for relevant visual description data. The best-fit visual description data is selected to generate the visual display, which then appears during the audio programme.

20 MPEG is the preferred delivery stream for the present invention. It can carry several video and audio streams. The decoder can decode and render two audio-visual streams simultaneously.

25 The exact types of applications vary, depending on the broadcast or network services and hardware capabilities of the receiver. In TV applications such as a music video, which already includes a video signal, the programme-associated data may be used to generate relevant video clips, images, graphics and textual display and on screen displays (particularly interactive ones) as a first video signal and superimposing or  
30 overlaying it onto the music video (the second video signal). However, there will also be applications where the display of visual description data generated is the only signal displayed.

Additionally, when a user plays an audio programme containing audio description  
35 data, an icon appears on a display, indicating that valid programme-associated data is

present. If the user presses a "Start Visual" button, the receiver searches for best-fit visual description data and generates the relevant visual display. By using pre-assigned remote control buttons, the user may navigate through interactive programs that are carried in the visual description data. An automatic option is also provided to  
5 start the best-fit visual display when incoming audio description data is detected.

The receiver is free to decide which visual description data shall be selected and how long each visual description data shall be played. Typically, search criteria are obtained from the audio description data when it is received. The visual description  
10 database is searched, based on the search criteria and a list of file locations is constructed, based on playing order. If the visual description play feature is enabled, this data is then played in this sequence. If another search criteria is obtained, the remaining visual description data is played out and the above procedure is followed to construct a new list of data matching the new criteria. User options are included to  
15 refine the cognitive algorithm and searching process. In the implementations, the visual description data may be declarative (e.g. HTML) or procedural (e.g. JAVA), depending on the set of Application Programming Interface functions available for the receiver.

20 Figure 3 is a schematic view of what happens at a receiver.

A digital television (DTV) source MPEG-2 stream 102 comprises visual description data 104, an encoded video stream 106 and an encoded audio stream 108 provides each stream, accessible separately. An MPEG-2 transport stream is preferred in DTV  
25 as it has robust error transmission. The visual description data is carried in an MPEG-2 private section. The encoded video stream is carried in MPEG-2 Packetised Elementary Stream (PES). The encoded audio stream also carries audio description data 110, which is separated out when the encoded audio stream is decoded.

30 Other sources 112, such as archives also provide second visual description data 114 and a second encoded video stream 116.

The two sets of visual description data and the two encoded video streams are provided to a search engine 118 as searchable material, whilst the audio description  
35 data is also input to the search engine as search information. Visual description data

that is selected is interpreted by a decoder to construct a video signal 120 (usually graphics or short video clips). It uses much less data to construct this video signal compared with the video stream. An encoded video signal that is selected is decoded to produce a second video signal 122.

5

In parallel, the decoding of the encoded audio stream, as well as providing audio description data 110 also provides audio signal 124.

10

A renderer 126 receives the two video signals and, because it is constructed in various layers (including graphics and OSD), is able to provide a combined video signal 128 in which multiple video signals overlap. The renderer also has an input from the audio description data. The combined video signal can be altered by a user select 130.

The audio signal is also rendered separately to produce sound 132.

15

An example of a format for the audio description data will now be described.

The audio description data is placed in an ancillary data section within each frame of an audio elementary stream. Table 1 shows the syntax of an audio frame as defined in ISO/IEC 11172-3 (MPEG – Audio).

20

Table 1: Syntax of audio frame

Syntax	No. of bits
frame() { header error_check audio_data() ancillary_data() }	32 16  no_of_ancillary_bits

25

The ancillary data is located at the end of each audio frame. The number of ancillary bits equals the available number of bits in an audio frame minus the number of bits used for header (32 bits), error check (16 bits) and audio. The numbers of audio data bits and ancillary data bits are both variable. Table 2 shows the syntax of the ancillary data used to carry the programme-associated data. The ancillary data is user definable, based on the definitions shown later, according to the audio content itself.

30

Table 2: Syntax of ancillary data

Syntax	No. of bits
<pre> ancillary_data() {     if ( (layer==1)    (layer==2) ) {         for (b=0; b&lt;no_of_ancillary_bits; b++) {             ancillary_bit         }     } } </pre>	1

5

The audio description data is created and inserted as ancillary data by the content creator or provider prior to distribution or broadcast.

Table 3 shows the syntax of the audio description data in each audio frame, residing in the ancillary data section.

10

Table 3: Syntax of audio description data

Syntax	No. of bits
<pre> audio_description_data() {     audio_description_identification     distribution_flag_bit     description_data_type     description_data_code     if (description_data_type == 0) {         audiovisual_pad_identification         audiovisual_clock_reference     }     else if (description_data_type &lt;= 15) {         user_data_code()     } } </pre>	13 1 5 5 16 16

15 The semantic definitions are:

audio\_description\_identification -- A 13-bit unique identification for user definable ancillary data carrying audio description information. It shall be used for checking the presence of audio description data relevant to the audio content.

distribution\_flag\_bit -- This 1-bit field indicates whether the following audio description data within the audio frame can be edited or removed. A '1' indicates no modification is allowed. A '0' indicates editing or removal of the following audio description data is possible for re-distribution or broadcast.

description\_data\_type -- This 5-bit field defines the type of data that follows. The data type definitions are tabulated in Table 4.

description\_data\_code -- This 5-bit field contains the predefined description code for description\_data\_type greater than 15. It is undefined for description\_data\_type between 0 to 15.

audiovisual\_pad\_identification -- A 16-bit programme-associated data identification for application where the audio content, including the audio description data, comes with optional associated visual description data. The receiver may look for matching visual description data having the same identification in the receiver's memory system.

audiovisual\_clock\_reference -- This 16-bit field provides a clock reference for the receiver to synchronise decoding of the visual description data. Each count is 20msec.

user\_data\_code -- User data in each audio frame to describe text characters and Karaoke text and timing information.

Table 4 shows the definitions of the description\_data\_type that defines the data type for description\_data\_code.

Table 4: Definitions of description\_data\_type

Value	Definitions	Data Loop
0	Identification followed by Clock Reference.	
1	Title description.	√
2	Singer/Group name description.	√
3	Music company name description.	√
4	Service provider description.	√
5	Service information description	√
6	Current event description	√
7	Next event description	√
8	General text description	√
9-12	Reserved	√
13	Karaoke text and timing description	√
14	Web-links	√



15	Reserved	√
16	Style	
17	Theme	
18	Events	
19	Objects	
20-31	Reserved	

A value of 0 indicates that the codes after `description_data_code` shall contain `audiovisual_pad_identification` and `audiovisual_clock_reference` data. The former provides a 16-bit unique identification for applications where the present audio content comes with optional associated visual description data having the same identification number. When the receiver detects this condition, it may look for matching visual description data having the same identification in its memory system. If no matching visual description data is found, the receiver may filter incoming streams for the matching visual description data. The `audiovisual_clock_reference` provides a 16-bit clock reference for the receiver to synchronise decoding of the visual description data. Each count is 20msec. With 16-bit clock reference and a resolution of 20msec per count, the maximum total time without overflow is 1310.72 sec, and shall be sufficient for each audio music or song duration.

Table 5, 6 and 7 list the descriptions of the pre-defined `description_data_code` for "style", "theme" and "events" data type respectively. The `description_data_type` and `description_data_code` shall be used as a basis for implementing cognitive and searching processes in the receiver for deducing the best-fit visual description data to generate the visual display. The selection of visual description data may be different even for the same audio elementary stream, as it is up to the receiver's cognitive and search engines' implementations. User options may be added to specify preferred categories of visual description data.

Table 5: Definitions of `description_data_code` for `description_data_type` equals "style"

Value	Definitions	Value	Definitions
0	Reserved	11	Latin
1	Children's	12	Music
2	Christian & Gospel	13	New Age
3	Classical	14	Opera
4	Country	15	Pop
5	Dance	16	Rap
6	Folk	17	Rock
7	Instrumental	18	Sentimental

8	International	19	Soul
9	Jazz	20	Soundtracks
10	Karaoke	21-31	Reserved

Table 6: Definitions of description\_data\_code for description\_data\_type equals "theme"

Value	Definitions	Value	Definitions
0	Reserved	11	Kids
1	Action and adventure	12	Leisure and entertainment
2	Art and architecture	13	Love and romance
3	Beach, wet and wild	14	Music and musical
4	Business	15	Outdoors and nature
5	Family	16	Science fiction and fantasy
6	Food and wine	17	Sports
7	Fun	18	Supermarket
8	Health and beauty	19	Teens
9	Home and garden	20	Travel
10	Horror and suspense	21-31	Reserved

Table 7: Definitions of description\_data\_code for description\_data\_type equals "events"

Value	Definitions	Value	Definitions
0	Reserved	6	National day
1	Birthday	7	New year's day
2	Children's day	8	Sales
3	Chinese new year	9	Sports events
4	Christmas day	10	Wedding day or anniversary
5	Festive Celebrations	11-23	Reserved

The audio description data may be used to describe text and the timing information in audio content for Karaoke application. Table 8 shows the syntax of the karaoke\_text\_timing\_information residing in the ancillary data section of the audio frame. Table 8 falls into "user\_data\_code" in Table 3. This happens when "description\_data\_type" = 13 in Table 4.

Table 8: Syntax of karaoke\_text\_timing\_description()

Syntax	No. of bits
karaoke_text_timing_description() { karaoke_clock_reference iso_639_language_code start_display_time audio_channel_format	 16 24 16 2

upper_text_length	6
for (i=0;i<upper_text_length;i++) { upper_text_code	16
}	
reserved	2
lower_text_length	6
for (i=0;i<lower_text_length;i++){ lower_text_code	16
}	
for (i=0;i<upper_text_length+1;i++){ upper_time_code	16
}	
for (i=0;i<lower_text_length+1;i++){ lower_time_code	16
}	
}	

Audio channel information is provided in Table 9

Table 9: Definitions of audio\_channel\_format

5

Value	Definitions
0	Use default audio settings.
1	Music at left channel. Vocal at right channel.
2	Music at right channel. Vocal at left channel.
3	Reserved.

The semantic definitions are:

- karaoke\_clock\_reference -- This 16-bit field provides a clock reference for the receiver to synchronise decoding of the Karaoke text and time codes. It is used to set the current decoding clock reference in the decoder. Each count is 20msec.
- iso\_639\_Language\_Code -- This 24-bit field contains 3 character ISO 639 language code. Each character is coded into 8 bits according to ISO 8859-1.
- start\_display\_time -- This 16-bit field specifies the time for displaying the two text rows. It is used with reference to the karaoke\_clock\_reference. Each count is 20msec.
- audio\_channel\_format -- This 2-bit field indicates the audio channel format for use in the receiver for setting the left and right output. See Table 9 for definitions.
- upper\_text\_length -- This 6-bit field specifies the number of text characters in the upper display row.

- upper\_text\_code -- The code defining the text characters in the upper display row (from 0 to 64).
- lower\_text\_length -- This 6-bit field specifies the number of text characters in the lower display row.
- 5 lower\_text\_code -- The code defining the text characters in the lower display row (from 0 to 64).
- upper\_time\_code -- This 16-bit field specifies the scrolling information of the individual text character in the upper display row. It is used with reference to the karaoke\_clock\_reference. Each count is 20msec.
- 10 lower\_time\_code -- This 16-bit field specifies the scrolling information of the individual text character in the lower display row. It is used with reference to the karaoke\_clock\_reference. Each count is 20msec.

The karaoke\_clock\_reference starts from count 0 at the beginning of each Karaoke song. For synchronisation of Karaoke text with audio, the audio description data encoder is responsible for updating the karaoke\_clock\_reference and setting start\_display\_time, upper\_time\_code and lower\_time\_code for each Karaoke song.

15

In the receiver, the timing for text display and scrolling is defined in the start\_display\_time, upper\_time\_code and lower\_time\_code fields. The receiver's Karaoke text decoder timer shall be updated to karaoke\_clock\_reference. When the decoder count matches start\_display\_time, the two rows of text shall be displayed without highlighting. The scrolling information is embedded in the upper\_time\_code and lower\_time\_code fields. They are used for highlighting the text character display to make the scrolling effect. For example, the decoder will use the difference between the upper\_time\_code[n] and upper\_time\_code[n+1] to determine the scroll speed for text character in the upper row at nth position. A pause in scrolling is done by inserting a space text character. At the end of scrolling in the lower row, the decoder remove the text display and the decoder process repeats with the next start\_display\_time.

20

25

30 With 16 bit time code and a resolution of 20msec per count, the maximum total time without overflow is 1310.72 sec or 21 minutes and 50.72sec. The specification does not restrict the display style of the decoder model. It is up the decoder implementation to use the start\_display\_time and the time code information for displaying and

highlighting the Karaoke text. This enables various hardwares with different capabilities and On-Screen-Display (OSD) features to perform Karaoke text decoding.

5 The visual description data may be in various formats, as mentioned earlier. This tends to be platform dependent. For example in MHP (Multimedia Home Platform) receivers, JAVA and HTML are supported.

10 In audio only applications, it may be desirable to insert programme-associated data to generate a relevant, exciting and interesting visual display for a listener. To generate such a visual display, a method of encoding and inserting the programme-associated data in the audio elementary streams, as well as a technique of decoding, interpreting and generating the visual display has been introduced.

15 Developing visual content relevant to the audio or TV programme requires significant resources. Getting the viewer to access these additional data service information is important for successful commercial implementations. In most cases, the viewer would find a TV programme uninteresting after having watched the programme and is less likely to be watching it many more times. However, for audio applications, the listener is more likely to repeat the same music and song over and over again. Thus, the  
20 solution of generating visual display relevant to the audio content includes the option of generating different displays to arouse the viewer's attention, even when playing the same audio content. To reduce the cost of content development for generating the visual display, the present invention enables sharing and reuse of the programme-associated data among different audio and TV applications.

25 In TV applications such as music video, the programme-associated data carried in the audio elementary stream may be used to generate relevant graphics and textual display on top of the video. Thus, one embodiment provides a method that enables additional visual content superimposing or overlaying onto the video.

30 The implementations are mainly software. Applications for editing audio description data can be used to assist the content creator or provider to insert relevant data in the audio elementary stream. Software development tools can be used to generate the visual description data for inserting in the transport or programme streams as private  
35 data. In the receiver, when the audio programme containing the audio description data

is played, the receiver extracts the audio description data and searches its memory system for relevant visual description data that have been extracted or downloaded previously. The user may also generate individual visual description data. The best-fit visual description data is selected to generate the visual display.

5

With current advances in technologies, especially in the area of digital TV, there are many opportunities to develop visual and interactive programmes on top of a background video. This invention provides an effective means of adding further information relevant to the audio programme. It creates an option for the content creator to insert or modify relevant descriptive information or links for generating relevant visual content prior distributing or broadcasting. The programme-associated data carried in the ancillary data section of the audio elementary stream provides general description of the preferred classification or categories for use by the decoder to generate relevant visual display and interactive applications. A commercially viable scheme that fits into digital audio and TV broadcasting, as well as other multimedia platforms is beneficial to content providers, broadcasters and consumers. Thus the invention can be used in multimedia applications such as in digital TV, digital audio broadcasting, as well as in the Internet domain, for distribution of programme-associated data for audio contents.

10  
15  
20

In terms of positioning the constructed visual description data, this can be placed as desired, for instance as is described in the co-pending patent application filed by the same applicant on 4 October 2002 and entitled Visual Contents in Karaoke Applications, the entire contents of which are herein incorporated by reference.

25

Although only single embodiments of an encoder and a receiver and of the audio description data have been described, other embodiments and formats can readily be used, falling within the scope of what has been invented, both as claimed and otherwise.

CLAIMS

1. A method of providing an audio signal with an associated video signal, comprising the steps of:
  - 5 decoding an encoded audio stream to provide an audio signal and audio description data; and
  - providing an associated first video signal at least part of whose content is selected according to said audio description data.
- 10 2. A method according to claim 1, further comprising the earlier step of encoding said audio signal and said audio description data into said encoded audio stream.
3. A method according to claim 1 or 2, further comprising the step of decoding a second video signal from an encoded video stream.
- 15 4. A method according to any one of the preceding claims, wherein said providing step comprises:
  - using said audio description data to select visual description data appropriate to the content of said audio signal;
  - 20 constructing video content from said selected visual description data; and
  - providing said first video signal including the constructed video content.
5. A method according to claim 4, further comprising the step of extracting said visual description data from a transport stream.
- 25 6. A method according to claim 5, wherein said visual description data is extracted from private data within said transport stream.
7. A method according to claim 5 or 6 when dependent on at least claim 3, wherein said transport stream further comprises said encoded video and audio streams.
- 30 8. A method according to claim 7, wherein said audio description data in said encoded audio stream includes identification data and clock reference data for use with said visual description data in said same transport stream.
- 35

9. A method according to claim 8, wherein descriptors corresponding to said identification data and clock reference data are stored in private sections of said visual description data.
- 5
10. A method according to any one of claims 7 to 9, wherein said audio stream, said video stream and said visual description data are multiplexed into said transport stream which is transmitted in a television signal.
- 10 11. A method according to any one of claims 7 to 10, wherein said step of using said audio description data to select appropriate visual description data comprises selecting visual description data from the same transport stream.
- 15 12. A method according to any one of claims 4 to 11, further comprising the step of storing said extracted visual description data.
13. A method according to claim 12 when not dependent on claim 11, wherein said step of using said audio description data to select appropriate visual description data comprises selecting stored visual description data.
- 20 14. A method according to any one of claims 4 to 13, further comprising the step, prior to the step of extracting said visual description data, of encoding said visual description data.
- 25 15. A method of delivering programme-associated data to generate relevant visual display for audio contents, said method comprising the steps of:  
encoding an audio signal and audio description data associated therewith into an encoded audio stream;  
encoding visual description data; and  
30 combining said encoded audio stream and said visual description data.
16. A method according to claim 15, wherein said visual description data can be combined into a first video signal.



17. A method according to claim 15 or 16, further comprising encoding a second video signal into an encoded video stream.
18. A method according to claim 17, further comprising combining said encoded video stream with said visual description data and said encoded audio stream into a transport stream.
19. A method according to claim 18, further comprising transmitting said transport stream in a television signal.
20. A method according to claim 18 or 19, wherein said visual description data does not relate to the encoded video signal in the same transport stream.
21. A method according to claim 18, 19 or 20, wherein said visual description data does not relate to the encoded audio signal in the same transport stream.
22. A method according to any one of claims 4 to 14 and 18 to 21, wherein said transport stream is an MPEG stream.
23. A method according to any one of claims 15 to 22 in combination with the method of any one of claims 1 to 14.
24. A method according to any one of claims 3 to 23, wherein said visual description data comprises one or more of the group comprising: video clips, still images, graphics and textual descriptions.
25. A method according to any one of claims 3 to 24, wherein said visual description data is classified for use with at least one of: at least one style of audio content, at least one theme of audio content and at least one type of event for which it might be suitable.
26. A method according to any one of the preceding claims, wherein said audio description data comprises data relating to at least one of the group comprising: singer identification, group identification, music company identification, service provider identification and karaoke text.

27. A method according to any one of the preceding claims, wherein said audio description data comprises data relating to the style of said audio signal.
- 5 28. A method according to any one of the preceding claims, wherein said audio description data comprises data relating to the theme of audio signal.
- 10 29. A method according to any one of the preceding claims, wherein said audio description data comprises data relating to the type of event for which said audio signal might be suitable.
- 15 30. A method according to any one of the preceding claims, wherein said audio description data is encoded within frames of said encoded audio stream, which frames also contain said audio signal.
31. A method according to claim 30, wherein said audio description data is encoded as ancillary data within audio frames of said audio stream.
- 20 32. Apparatus for providing an audio signal with an associated video signal, comprising:  
audio decoding means for decoding an encoded audio stream to provide an audio signal and audio description data; and  
first video signal means for providing an associated first video signal at least part of whose content is selected according to said audio description data.
- 25 33. Apparatus according to claim 32, further comprising video decoding means for decoding a second video signal from an encoded video stream.
- 30 34. Apparatus according to claim 32 or 33, wherein said first signal means comprises:  
selecting means for using said audio description data to select visual description data appropriate to the content of said audio signal;  
constructing means for constructing video content from said selected visual description data; and

means for providing said first video signal including the constructed video content.

- 5 35. A method according to claim 34, further comprising extracting means for extracting said visual description data from a transport stream.
36. Apparatus according to claim 35, wherein said extracting means is operable to extract said visual description data from private data within said transport stream.
- 10 37. Apparatus according to claim 35 or 36 when dependent on at least claim 32, operable when said transport stream further comprises said encoded video and audio streams.
- 15 38. Apparatus according to claim 37, operable when said audio description data in said encoded audio stream includes identification data and clock reference data for use with said visual description data in said same transport stream.
- 20 39. Apparatus according to claim 38, operable when descriptors corresponding to said identification data and clock reference data are stored in private sections of said visual description data.
- 25 40. Apparatus according to any one of claims 37 to 39, operable when said audio stream, said video stream and said visual description data are multiplexed into said transport stream which is transmitted in a television signal.
41. Apparatus according to any one of claims 37 to 40, wherein said selecting means is operable to select appropriate from the same transport stream as the visual description data.
- 30 42. Apparatus according to any one of claims 35 to 41, further comprising storing means for storing said extracted visual description data.
- 35 43. Apparatus according to claim 42, wherein said selecting means is operable to select appropriate visual description data from the storing means.

44. A system for delivering programme-associated data to generate relevant visual display for audio contents, comprising:  
audio encoding means for encoding an audio signal and audio description data associated therewith into an encoded audio stream;  
5 description data encoding means for encoding visual description data; and  
combining means for combining said encoded audio stream and said visual description data.
45. A system according to claim 44, further comprising video encoding means for  
10 encoding a second video signal into an encoded video stream.
46. A system according to claim 45, wherein said combining means is operable to combine said visual description data, said encoded audio stream and said encoded video stream into a transport stream.  
15
47. A system according to claim 46, wherein said combining means is operable to combine said visual description data with encoded video signal to which it does not relate, in the same transport stream.
- 20 48. A system according to claim 46 or 47, wherein said combining means is operable to combine said visual description data with encoded audio signal to which it does not relate, in the same transport stream.
- 25 49. A system according to any one of claims 46 to 48 or apparatus according to any one of claims 35 to 43, wherein said transport stream is an MPEG stream.
50. A system according to any one of claims 44 to 50 in combination with the apparatus of any one of claims 31 to 43.
- 30 51. A system according to any one of claims 44 to 50 or apparatus according to any one of claims 31 to 43 and 50, wherein said visual description data comprises one or more of the group comprising: video clips, still images, graphics and textual descriptions.

52. A system according to any one of claims 44 to 51 or apparatus according to any one of claims 31 to 43, 50 and 51, wherein said visual description data is classified for use with at least one of: at least one style of audio content, at least one theme of audio content and at least one type of event for which it might be suitable.
53. A system according to any one of claims 44 to 52 or apparatus according to any one of claims 31 to 43 and 50 to 52, wherein said audio description data comprises data relating to at least one of the group comprising: singer identification, group identification, music company identification, service provider identification and karaoke text.
54. A system according to any one of claims 44 to 53 or apparatus according to any one of claims 31 to 43 and 50 to 53, wherein said audio description data comprises data relating to the style of said audio signal.
55. A system according to any one of claims 44 to 54 or apparatus according to any one of claims 31 to 43 and 50 to 54, wherein said audio description data comprises data relating to the theme of audio signal.
56. A system according to any one of claims 44 to 55 or apparatus according to any one of claims 31 to 43 and 50 to 55, wherein said audio description data comprises data relating to the type of event for which said audio signal might be suitable.
57. A system according to any one of claims 44 to 56 or apparatus according to any one of claims 31 to 43 and 50 to 56, wherein said audio encoding means is operable to encode said audio description data within frames of said encoded audio stream, which frames also contain said audio signal.
58. A system or apparatus according to claim 57, wherein said audio encoding means is operable to encode said audio description data as ancillary data within audio frames of said audio stream.

59. A method of delivering programme-associated data to generate relevant visual display for audio contents, said method, comprising:  
encoding audio description data relevant to the audio contents in one or more audio elementary streams; and  
5 encoding visual description data created for audio contents for generating a visual display; wherein  
said visual description data is relevant to at least one of the groups comprising:  
a generic audio style, a generic audio theme, special events and specific objects.
- 10 60. The method of claim 59, further comprising the preceding steps of:  
specifying preferred visual displays for the frames of said audio elementary stream; and  
constructing said audio description data using information relating to said preferred visual displays.
- 15 61. The method of claim 58, wherein said specifying step comprises identifying at least one of:  
the style of the audio content;  
the theme of said audio frame;  
20 an event associated with said audio frame; and  
keywords in any lyrics of said audio frame;  
and further comprising specifying a most preferred visual display after the identifying step.
- 25 62. The method of claim 60 or 61, wherein said specifying step comprises specifying the preferred visual display for each of said frames.
63. The method of any one of claims 59 to 62, further comprising inserting said audio description data in ancillary data sections of said audio frames in said audio elementary stream.
- 30 64. The method of any one of claims 59 to 63, wherein said constructing step comprises:  
specifying a unique identification code;  
35 specifying a distribution flag for indicating distribution rights;

specifying the data type;  
inserting text description describing the audio content;  
inserting data code describing said preferred visual display; and  
inserting user data code for generating the visual display.

5

65. The method of any one of claims 59 to 64, further comprising:  
encoding background video into a video elementary stream; and  
encoding the audio contents into said one or more audio elementary streams;  
and wherein said audio description data describes said audio contents.

10

66. The method of any one of claims 59 to 65, wherein the step of encoding visual  
description data comprises encoding the visual description data into private data to  
be carried in a transport stream.

15

67. The method of claims 65 and 66, further comprising multiplexing said video  
elementary stream, said one or more audio elementary streams and said private  
data into a transport stream for broadcast.

20

68. The method of any one of claims 59 to 67, further comprising delivering said  
audio description data and said video description data to a receiver for decoding  
and for generating said visual display.

25

69. The method of any one of claims 59 to 68, further comprising the step of  
providing said visual description data by downloading it from external media or  
creating it at a user terminal.

30

70. A method of delivering Karaoke text and timing information to generate a  
Karaoke visual display for an audio song, said method comprising:  
encoding said audio song into an audio elementary stream;  
inserting clock references for use in synchronising decoding of said Karaoke  
text and timing information with said audio song in said audio elementary stream;  
inserting channel information of said audio song in said audio elementary  
stream;

35

- inserting said Karaoke text information for said audio song in said audio  
elementary stream; and

inserting said Karaoke timing information for generating scrolling said Karaoke text in said audio elementary stream.

71. The method of any one of claims 1 to 31 and 59 to 70 being used in digital TV broadcast and or reception.
72. Apparatus for generating relevant visual display for audio contents, comprising:  
storing means for storing visual description data that generate the visual display;  
playing means for playing said audio contents carried in an audio elementary stream;  
extracting means for extracting audio description data for said audio contents from said audio elementary stream;  
selecting means for selecting preferred visual description data from said storing means using information from said audio description data; and  
executing means for executing said visual description data to generate said visual display.
73. Apparatus according to claim 72, wherein said executing means is operable to execute interactive programmes carried in said visual description data.
74. Apparatus according to claim 72 or 73, further comprising:  
receiving means for receiving a multiplexed transport stream containing one or more of said audio elementary streams and said visual description data carried as private data.
75. A system for connecting audio and visual contents, comprising:  
downloading means for downloading audio elementary streams for said audio contents and for downloading visual description data;  
creating and editing means for creating and editing audio description data relevant to said audio contents carried in said audio elementary streams and for creating and editing visual description data for generating said visual contents;  
selecting means for selecting said visual description data that best fits the audio description data for generating a visual display;  
user operable means for modifying the behaviour of said selecting means; and



processor means for executing said visual description data to generate the display.

- 5 76. A system according to claim 75, wherein said selecting means comprise cognitive and search engines.
77. A system according to claim 75 or 76, being a home entertainment system.
- 10 78. A method of providing an audio signal with an associated video signal substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.
- 15 79. A method of delivering programme-associated data to generate relevant visual display for audio contents substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.
- 20 80. Apparatus for providing an audio signal with an associated video signal constructed and arranged to operate substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.
81. A system for providing an audio signal with an associated video signal constructed and arranged to operate substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.
- 25 82. A system for delivering programme-associated data to generate relevant visual display for audio contents constructed and arranged to operate substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.
- 30 83. Apparatus according to any one of claims 32 to 43, 51 to 58, 72 to 74 and 80 or a system according to any one of claims 44 to 58, 75 to 77, 81 and 82, operable according to the method of any one of claims 1 to 31, 59 to 71, 78 and 79.

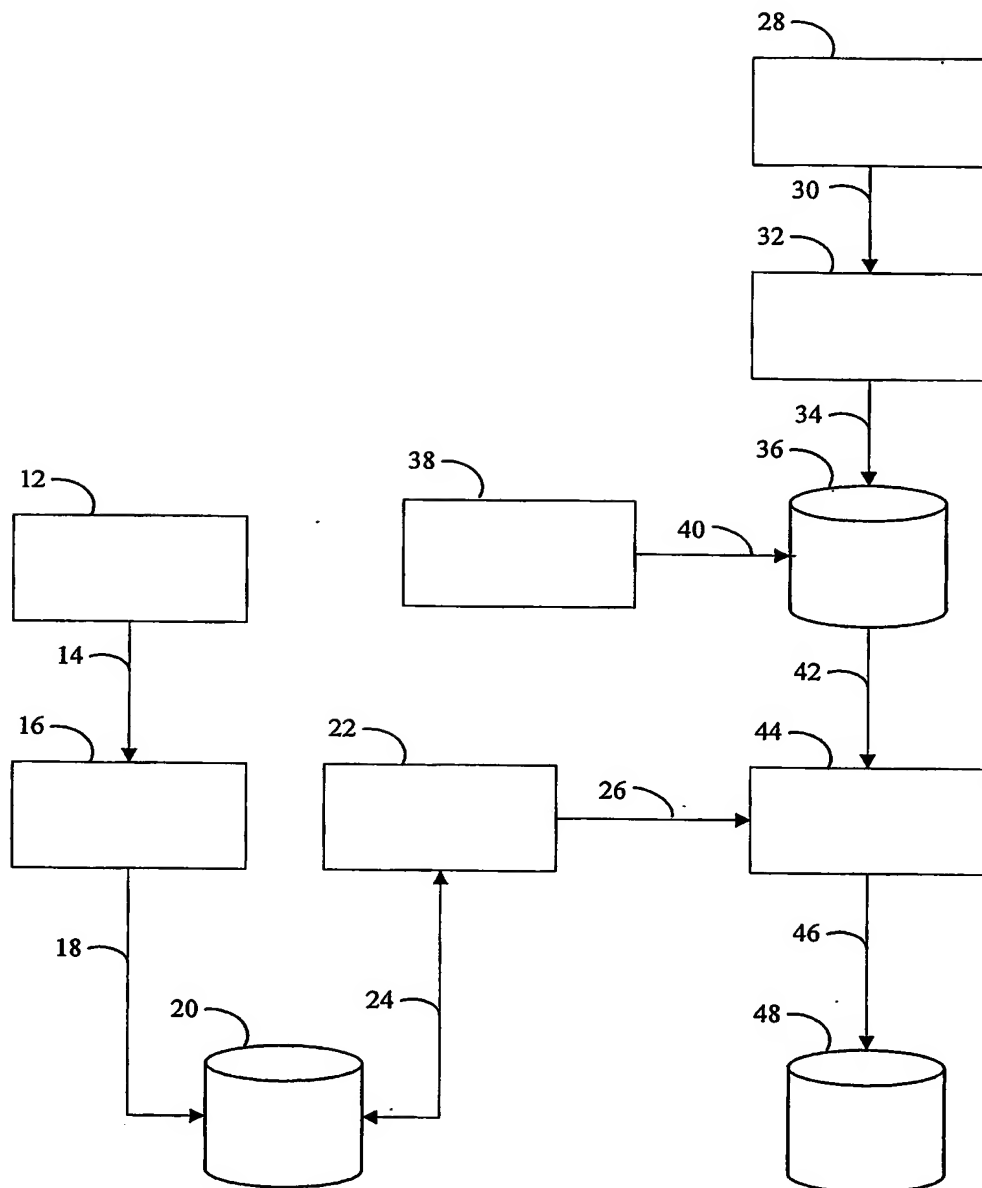


Figure 1

2/3

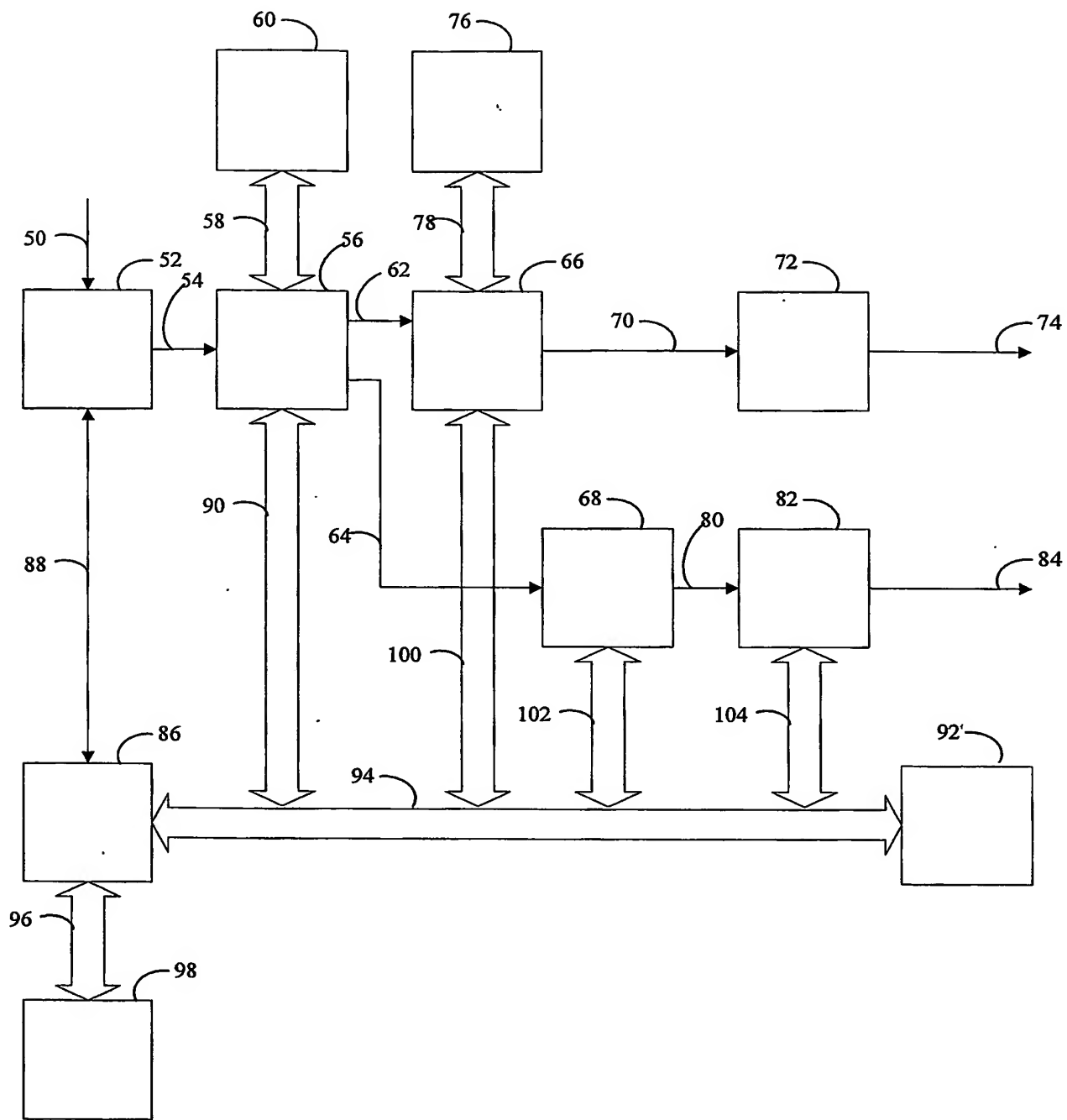


Figure 2

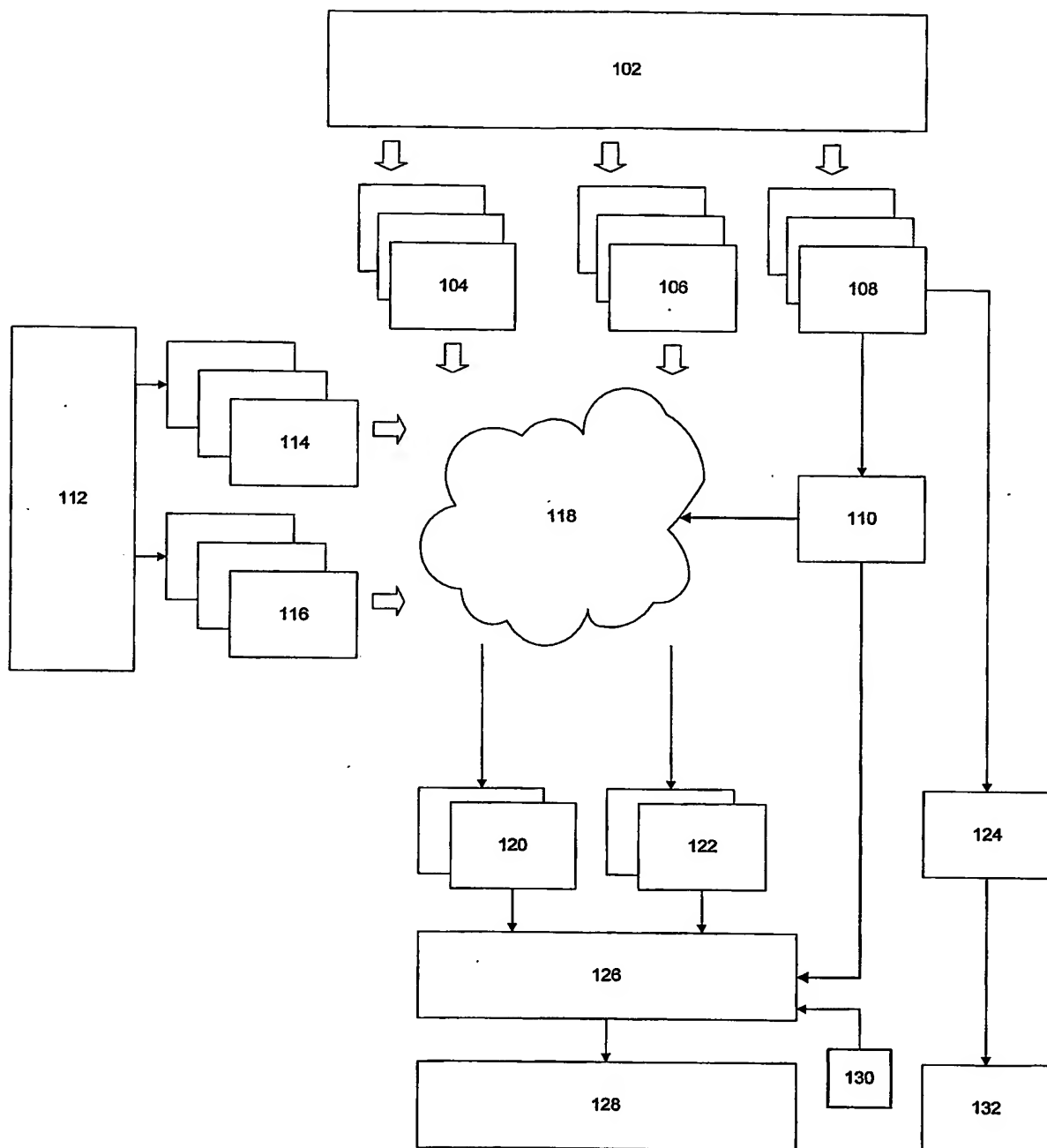


Figure 3

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG03/00233

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>												
Int. Cl. <sup>7</sup> : G06F 17/00												
According to International Patent Classification (IPC) or to both national classification and IPC												
<b>B. FIELDS SEARCHED</b>												
Minimum documentation searched (classification system followed by classification symbols)												
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched												
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) USPTO, DWPI, internet (mp3, mp3i, audio, music, id3, visualization, video, etc.)												
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>												
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.										
X, P	WO 02/103484 A2 (FIRST INTERNATIONAL DIGITAL, INC.), 27 December 2002 the whole document	1-77										
X	WO 02/071021 A1 (FIRST INTERNATIONAL DIGITAL, INC.), 12 September 2002 the whole document (in particular, paragraphs 81-83)	1-3, 26-33, 71, 83										
X	WO 01/61684 A1 (FIRST INTERNATIONAL DIGITAL, INC.), 23 August 2001 the whole document	70										
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C <input checked="" type="checkbox"/> See patent family annex												
<p>* Special categories of cited documents:</p> <table border="0"> <tr> <td>"A" document defining the general state of the art which is not considered to be of particular relevance</td> <td>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>"E" earlier application or patent but published on or after the international filing date</td> <td>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>"O" document referring to an oral disclosure, use, exhibition or other means</td> <td>"&amp;" document member of the same patent family</td> </tr> <tr> <td>"P" document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	"P" document published prior to the international filing date but later than the priority date claimed	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention											
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone											
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art											
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family											
"P" document published prior to the international filing date but later than the priority date claimed												
Date of the actual completion of the international search 18 November 2003		Date of mailing of the international search report 19 NOV 2003										
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustalia.gov.au Facsimile No. (02) 6285 3929		Authorized officer  <b>M. D. HOLLINGWORTH</b> Telephone No : (02) 6283 2024										

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG03/00233

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,369,822 B1 (PEEVERS et al), 9 April 2002 the whole document	1-83
A	US 6,395,969 B1 (FUHRER), 28 May 2002 the whole document	1-83
O	<i>MP3i Creator: Features</i> , mp3icreator.com website, 4 October 2002 <URL: <a href="http://web.archive.org/web/20021004095609/http://www.mp3icreator.com/creator/features/">http://web.archive.org/web/20021004095609/http://www.mp3icreator.com/creator/features/</a> >	1-83

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG03/00233

**Box I** Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos :  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos :  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos :  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

**Box II** Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG03/00233

**Supplemental Box**

(To be used when the space in any of Boxes I to VIII is not sufficient)

**Continuation of Box No: II**

The international application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept. In coming to this conclusion, the International Searching Authority has found that there are two inventions:

1. Claims 1-69 and 71-83, directed to systems and methods for providing visual content for an audio stream. The selection of the visual content according to audio description data encoded in the stream is considered to be a first "special technical feature."
2. Claim 70, directed to a method of delivering karaoke text and timing information. The insertion of karaoke text and timing information into an audio stream is considered to be a second "special technical feature."

Although these groups of claims both feature the encoding of non-audio data into an audio stream, this is well known and cannot constitute a "special technical feature" uniting the claims. Since the claims share no other technical features, a "technical relationship" between the inventions, as define in PCT rule 13.2, does not exist. Accordingly, the international application does not relate to one invention or to a single inventive concept.



**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
**PCT/SG03/00233**

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member	
WO	02103484	US	2002193895
WO	02071021	US	2002165720
WO	0161684	AU	17654/01
US	6369822	NONE	US 6442517
US	6395969	NONE	
END OF ANNEX			